entities are reported as such. The advertisement sampling system uses intelligent agent technology to retrieve Web pages at various frequencies to obtain a representative sample. This allows the Cloudprober to accurately assess how frequently each advertisement appears in the traffic data. After the Cloudprober fetches a Web page, the advertisement sampling system extracts the advertisements from the Web page. In the preferred embodiment, the advertisement extractor, also known as the "extractor", invokes an automatic advertisement detection ("AAD") process, a heuristic extraction process, to automatically extract all of the advertisements from the Web page.

Replace the paragraph on page 6, line 9 through page 6, line 10 with the following:

Figure 7D is a flow diagram that describes, in greater detail, the process of probing the Internet to gather sample data from Figure 7A.

Replace the paragraph on page 8, line 10 through page 8, line 13 with the following:

2. "Client-Side Panel Collection" retrieves sample data from each panelist via a client-side mechanism and transfers that data to a collection repository. The client-side mechanism may monitor the browser location bar, user browser, a client-side proxy, or TCP/IP stack hooks.

Replace the paragraph on page 9, line 9 through page 9, line 18 with the following:

The traffic analysis system 210 receives raw traffic data from the traffic sampling system 120. The traffic analysis system 210 cleanses the raw traffic data by removing information from the traffic data that may identify a particular user on the Internet 100 and then stores the anonymous data in the database 200. The traffic analysis system 210 estimates the global traffic to every significant Web site on the Internet 100. The present invention uses this data not only for computing the

number of advertising impressions given an estimate of the frequency of rotation on that page, but also in the probe mapping system 320. In one embodiment, the traffic analysis system 210 receives traffic data from a cache site on the Internet 100. The goal is to accurately measure the number of page views by individual users, and therefore the number of advertising impressions.

Replace the paragraph on page 9, line 19 through page 10, line 6 with the following:

The advertisement sampling system 220 uses the anonymous traffic data to determine which URLs to include in the sample retrieved from the Web server 112. The advertisement sampling system 220 contacts the Web server 112 through the Internet 100 to retrieve a URL 114, 116, 118 and extract the advertisements therein along with the accompanying characteristics that describe the advertisements. The success rate for retrieval of creatives is high. Analysis indicates that the present invention captures over 95% of creatives served. The advertisement sampling system 220 stores these advertisement characteristics in the database 200. The advertisement sampling system 220, for example, the Online Media Network Intelligent Agent Collection ("OMNIAC"), or the Cloudprober, repeatedly probes prominent Web sites, extracts advertisements from each Web page returned by the probe, and classifies the advertisements in each Web page by type, technology and advertiser.

Replace the paragraph on page 10, line 7 through page 10, line 10 with the following:

The traffic analysis system 210 and the advertisement sampling system 220 also present the data retrieved from the Internet 100 to the statistical summarization system 230 for periodic

processing. The statistical summarization system 230 calculates the advertising frequency, impressions, and spending on a per site and per week basis.

Replace the paragraph on page 11, line 1 through page 11, line 2 with the following:

The traffic analysis system 210 includes an anonymity system 310 and traffic summarization 312 process.

Replace the paragraph on page 12, line 3 through page 12, line 9 with the following:

The Web page retrieval system 322 uses the probe map generated by the probe mapping system 320 to determine which Web pages it needs to sample and the frequency of the sampling. For each URL in the probe map generated by the probe mapping system 320, the Web page retrieval system 322 fetches a Web page, extracts each advertisement from the Web page, and stores the advertisement's attributes in the database 200. The data retrieved from each URL in the probe map is used to calculate the frequency with which each advertisement is shown on a particular Web site

Replace the paragraph on page 14, line 9 through page 15, line 5 with the following:

The structural classifier 328 performs structural fragment analysis on the XML representation of the Web page by determining the "physical type" of the fragment (i.e., the HTML source code used to construct the advertisement). Physical types that the present invention recognizes include banner, form, single link, and embedded content. Banner advertisement fragments include a single HTML link having one or two enclosed images and no FORM or IFRAME tag. Form advertisement fragments include a single HTML form having no IFRAME tag. Single link advertisement fragments include a link with textual, but no IMG, FORM, or IFRAME tags. Embedded content

advertisement fragments reference an external entity using an IFRAME tag. After performing this analysis, the structural classifier 328 updates the advertisement fragment in the database. For a banner advertisement fragment, the structural classifier 328 stores the link and image URL's in the database 200. A form advertisement fragment requires the creation of a URL by simulating a user submission that sets each HTML control to its default value. The structural classifier 328 stores this URL and the "form signature" (i.e., a string that uniquely describes the content of all controls in the form) in the database 200. For a single text advertisement fragment, the structural classifier 328 stores the URL for the link and all text contained within the link in the database 200. For embedded content advertisement fragments, the structural classifier 328 stores the URL associated with the external reference in the database 200. This URL is loaded by the system, and the referenced document is loaded. Once the loaded document has been structurally analyzed, the original fragment inherits any attributes that result from analysis of the new fragment.

Replace the paragraph on page 16, line 19 through page 17, line 14 with the following:

The operator 262 uses the site administration 342 module of the user interface 240 to simplify the administration of the site definitions. Analysts from the Internet Advertising Bureau estimate that over 90% of all Web advertising dollars are spent on the top fifty Web sites. Site selection begins by choosing the top 100 advertisements by considering data from Media Metrix, Neilsen/Net Ratings, and the proxy traffic data in the database 200. These lists are periodically updated to demote Web sites with low traffic levels and promote new sites with high traffic levels. The present invention also includes Web sites that provide significant content in key industries. A

site chosen for inclusion in the site definitions must have the structure of the site analyzed to remove sections that do not serve advertisements, originate from foreign countries, or are part of a frame set. Sites that originate from a foreign country, such as yahoo.co.jp, sell advertising in the host country, and therefore are not applicable to the measurements calculated by the present invention. Web sites that use an HTML frameset are treated very carefully to only apply rotation rates to the traffic from the sections of the frameset that contain the advertisement. These combined exclusions are key to making accurate estimates of advertising impressions. The present invention also tags sections that cannot be measured directly, due to registration requirements (e.g., mail pages). Since Web sites change frequency, this structural analysis is repeated periodically. Eventually the analysis stage will automatically flag altered sites to allow even more timely updates.

Replace the paragraph on page 17, line 15 through page 17, line 23 with the following:

The media editor 264 uses the taxonomy administration 344, advertising content classification 346, and rate card collection 348 modules of the user interface 240. The taxonomy administration 344 module simplifies the creation and maintenance of the attributes assigned to advertisements during content classification including the advertisement's industry, company, and products. The taxonomy names each attribute and specifies its type, ancestry and segment membership. For example, a company Honda, might be parented by the Automotive industry and belong to the industry segment Automotive Manufactures. The advertising content classification 346 component assists the media editor 264 with performing the content classification.

Replace the paragraph on page 18, line 1 through page 18, line 16 with the following:

The structural classifier 328 performs automated advertisable assignment to determine what the advertisement is advertising. This process includes assigning "advertisables" (i.e., attributes describing each "thing" that the advertisement is advertising) to each advertisement fragment. In another embodiment of the present invention, the advertisement sampling system 220 uses an extensible set of heuristics to assign advertisables to each advertisement. In the preferred embodiment, however, the only automatic method employed is location classification. Location classification relies on the destination URL in order to assign a set of advertisables to an advertisement. A media editor 264 uses the user interface 240 to maintain the set of classified locations. For example, the first time a media editor observes an advertisement in which the click-thru URL is www.honda.com, he can enter this URL as pertaining to the advertiser "Honda Motors". Any subsequent advertisement that includes the same click-thru URL will also be recognized as a Honda advertisement. A classified location comprises a host, URL path prefix, and set of advertisables. Location classification assigns a classified location advertisable to an advertisement if the host in the destination URL matches the host of the classified location and the path prefix in the classified location matches the beginning of the path in the destination URL.

Replace the paragraph on page 18, line 17 through page 18, line 22 with the following:

The structural classifier 328 performs human advertisable assignment and verification as a quality check of the advertisable data. This phase is the most human intensive. A media editor 264 uses a graphical user interface module in the user interface 240 to display each advertisement, verify

automatic advertisable assignments, and assign any other advertisables that appear appropriate after inspection of the advertisement and the destination of the advertisement. The location classification database is also typically maintained at this time.

Replace the paragraph on page 19, line 1 through page 19, line 8 with the following:

The media editor 264 uses the rate card collection 348 module to enter the contact and rate card information for a Web site identified by the traffic analysis system 210, as well as, designated advertisers. Rate card entry includes the applicable quarter (e.g., Q4 2000), advertisement dimensions in pixels, fee structure (e.g., CPM, flat fee, or per click), cost schedule for buys of various quantities and duration. The media editor 264 also records the URL address of the online media kit and whether rates are published therein. Contact information for a Web site or advertiser includes the homepage, name, phone and facsimile numbers, email address, and street address.

Replace the paragraph on page 19, line 18 through page 20, line 11 with the following:

The first step in the process is to normalize the results from the traffic analysis system 210. The traffic analysis system 210 provides the traffic received by each Web page in the traffic data sample. Figure 4A depicts the exemplary traffic received at each Web page 411-416, 421-424 in the Internet 100 with the label "Traffic =". The probe map generated by the probe mapping system 320 includes an entry for each Web page 411-416, 421-424. The probe map also includes an "area" that each Web page 411-416, 421-424 consumes in the probe map. Figure 4A depicts the exemplary area that each Web page 411-416, 421-424 consumes in the probe map with the label "Area =". The normalized results are calculated by dividing the area that a Web page consumes in the probe map

by the sum of the area for each Web page in the traffic sample. In Figure 4A, the normalized value, or chance, for Web page P1 411 is the area for Web page P1 (i.e., 15) divided by the sum of the area for Web page P1, P2, P3, P4, P5, P6, Q1, Q2, Q3, and Q4 (i.e., 120). The normalized value is, therefore, 0.125, or 12.5%. In addition to the normalized value, the system also determines the scale by dividing the traffic for a Web page by the area for the Web page. In Figure 4A, the scale for Web page P1 411 is the traffic for Web page P1 (i.e., 150) divided by the area for Web page P1 (i.e., 15), therefore, the scale for Web page P1 is 10. Table 1 summarizes the scale and chance values for the remaining Web page in Figure 4A.

Replace the paragraph on page 21, line 2 through page 21, line 6 with the following:

Figure 4B depicts the exemplary Web page fetches at each Web page 411-416, 421-424 in the Internet 100 with the label "Fetches =". Figure 4B also depicts the exemplary number of views of each advertisement on a Web page 411-416, 421-424 with a label such as "A1 Views =" to indicate the number of views of advertisement A1, "A2 Views =" to indicate the number of views of advertisement A2, etc.

Replace the paragraph on page 21, line 7 through page 21, line 17 with the following:

Figure 4C depicts the exemplary Web page weighted fetches at each Web page 411-416, 421-424 in the Internet 100 with the label "Fetches =". Figure 4C also depicts the exemplary number of views of each advertisement on a Web page 411-416, 421-424 with a label such as "A1 Views =" to indicate the number of views of advertisement A1, "A2 Views =" to indicate the number of views of advertisement A2, etc. The next step in the calculation process is to calculate the Scaled

Fetches for each Web site 410, 420 by summing the product of the observed fetches from Figure 4B and the scale from Figure 4A, for each Web page 411-416, 421-424 in the Web site. Next, the calculation computes the Traffic for each Web site 410, 420 by summing the traffic from Figure 4A for each Web page 411-416, 421-424 in the Web site. The rate card, or CPM, is a value assigned by the media editor 264 for each Web site 410, 420. Table 2 summarizes the Scaled Fetches, Traffic, and CPM for Figures 4A through 4C.

Replace the paragraph on page 22, line 1 through page 22, line 13 with the following:

The next in the calculation process is to compute the Scaled Observations for each advertisement on each Web site 410, 420 by summing the product of the advertisement views from Figure 4B and the scale from Figure 4A, for each Web page 411-416, 421-424 in the Web site 410, 420. The final step in the calculation is to compute the advertising prevalence statistics (i.e., Frequency, Impressions, and Spending) for each advertisement in each Web site 410, 420. Frequency is computed by dividing the scaled observations by the scaled fetches for each advertisement in each Web site 410, 420. Impressions is computed by multiplying the Frequency by the Traffic from Table 2 above for each advertisement in each Web site 410, 420. Spending is computed by multiplying the Impressions by the CPM from Table 2 above for each advertisement in each Web site 410, 420. Table 3 summarizes the Scaled Observations, Frequency, Impressions, and Spending for Web site P 410 using the data in Figures 4A through 4C. Table 4 summarizes the Scaled Observations, Frequency, Impressions, and Spending for Web site Q 420 using the data in Figures 4A through 4C.

Replace the paragraph on page 25, line 15 through page 26, line 7 with the following:

Figure 5 illustrates a database structure that the advertising prevalence system 130 may use to store information retrieved by the traffic sampling system 120 and the Web page retrieval system 322. The preferred embodiment segments the database 200 into partitions. Each partition can perform functions similar to an independent database such as the database 200. In addition, a partitioned database simplifies the administration of the data in the partition. Even though the preferred embodiment uses database partitions, the present invention contemplates consolidation of these partitions into a single database, as well as making each partition an independent database and distributing each database to a separate general purpose computer workstation or server. The partitions for the database 200 of the present invention include sampling records 510, probing definitions 520, advertising support data 530, and advertising summary 540. The preferred embodiment of the present invention uses a relational database management system, such as the Oracle8*i* product by Oracle Corporation, to create and manage the database and partitions. Even though the preferred embodiment uses a relational database, the present invention contemplates the use of other database architectures such as an object-oriented database management system.

Replace the paragraph on page 28, line 15 through page 28, line 23 with the following:

If the site definition for "somesite" includes the inclusive URL prefix "com.somesite/" and the exclusive URL prefix "com.somesite/foo/bar", the application of this site definition to the above sample URLs listed above yields a system that includes URL 1, 2, and 4. URL 3 is not part of the site definition due to the explicit exclusion of "com.somesite/foo/bar". URL 5 is not part of

the site definition because it was never included in the inclusive URL prefix "com.somesite/". The user interface 240 populates the site definition 522 area in database 200. The probe mapping system 320 accesses the data in the site definition 522 area to determine which URLs to probe. The statistical summarization system 230 accesses the data in the site definition 522 area to determine traffic levels to sites by summing traffic to URLs included in a site.

Replace the paragraph on page 29, line 16 through page 30, line 8 with the following:

The advertisement extraction rule definition 526 area describes Extensible Markup Language ("XML") tags, typically representing a normalized HTML document, that indicate those portions of the content that the system considers to be advertisements. The system defines an extraction rule in terms of "XML structure" and "XML features". "XML structure" refers to the positioning of various XML nodes relative to others XML nodes. For example, an anchor ("A") node containing an image ("IMG") node is likely an advertisement. After using this structural detection process to match the advertisement content, the system examines the features of the content to determine if the content is an advertisement. To continue the previous example, if the image node contains a link ("href") feature that contains the sub-string "adserver", it is very likely an advertisement. Features may match based on a simple sub-string, as in the example, or a more complicated regular expression. Another form of extraction rule may point to a specific node in an XML structure using some form of XML path specification, such as a "Xpointer". The media editor 264 populates the advertisement extraction rule definition 526 area in the database 200. The advertisement extractor

326 of the advertisement sampling system 220 accesses the data in the advertisement extraction rule definition 526 area to determine which portions of each probed page represent an advertisement.

Replace the paragraph on page 30, line 22 through page 31, line 3 with the following:

The advertising information 534 area contains the data that describe what each unique advertisement recorded by the system advertises. The tables in the advertising information 534 area associate advertisables with advertisements. For example, the system may associate a company type of advertisable with a specific advertisement to indicate that the advertisement is advertising the company. The system uses the following methods to associate an advertisable with an advertisement:

Replace the paragraph on page 32, line 19 through page 32, line 21 with the following:

2. The number of impressions that an advertisement received. The system determines this statistic by measuring traffic levels for the Web site using the site definition and traffic data, and multiplying that measurement by the proportion of page views calculated above.

Replace the paragraph on page 35, line 3 through page 35, line 13 with the following:

The database objects comprising the "core schema" are most frequently used by various components of the OMNIAC system. Code bases that rely on this schema include implementation of the back-end processes that pull advertisements from the Web. Additionally, database schemas utilized by other components associated with OMNIAC are composed of some or all of the tables in the core schema. The core schema is conceptually composed of four sub-schemas including advertising, advertisements, probing, and sites. The advertising sub-schema holds information about

"advertiseable" entities along with which entities each advertisement is advertising. The advertisements sub-schema describes the advertisements that the system has located and analyzed. The probing sub-schema defines "when", "where", and "how" for the probing process. The sites sub-schema describes Web sites, including structural site definitions and rate card information.

Replace the paragraph on page 41, line 11 through page 42, line 8 with the following:

The presentation tier 620 retains the programs that manage the interface between the advertising prevalence system 130 and the client 140, account manager 260, operator 262, and media editor 264. In Figure 6, the presentation tier 620 includes the TCP/IP interface 622, the Web front end 624, and the user interface 626. A suitable implementation of the presentation tier 620 may use Java servlets to interact with the client 140, account manager 260, operator 262, and media editor 264 of the present invention via the hypertext transfer protocol ("HTTP"). The Java servlets run within a request/response server that handles request messages from the client 140, account manager 260, operator 262, and media editor 264 and return response messages to the client 140, account manager 260, operator 262, and media editor 264. A Java servlet is a Java program that runs within a Web server environment. A Java servlet takes a request as input, parses the data, performs logic operations, and issues a response back to the client 140, account manager 260, operator 262, and media editor 264. The Java runtime platform pools the Java servlets to simultaneously service many requests. A TCP/IP interface 622 that uses Java servlets functions as a Web server that communicates with the client 140, account manager 260, operator 262, and media editor 264 using the HTTP protocol. The TCP/IP interface 622 accepts HTTP requests from the client 140, account

manager 260, operator 262, and media editor 264 and passes the information in the request to the visit object 642 in the business logic tier 640. Visit object 642 passes result information returned from the business logic tier 640 to the TCP/IP interface 622. The TCP/IP interface 622 sends these results back to the client 140, account manager 260, operator 262, and media editor 264 in an HTTP response. The TCP/IP interface 622 exchanges data with the Internet 100 via the TCP/IP network adapter 614.

Replace the paragraph on page 42, line 9 through page 42, line 13 with the following:

The infrastructure objects partition 630 retains the programs that perform administrative and system functions on behalf of the business logic tier 640. The infrastructure objects partition 630 includes the operating system 636, and an object oriented software program component for the database management system ("DBMS") interface 632, administrator interface 634, and Java runtime platform 638.

Replace the paragraph on page 43, line 15 through page 44, line 5 with the following:

After the traffic analysis application 652 processes a URL 114, 116, 118 identified by the traffic sampling system 120, the visit object 642 invokes a method in the advertising sampling application 654 to retrieve the URL 114, 116, 118 from the Web site 110. The advertising sampling application 654 processes the retrieved Web page by extracting embedded advertisements and classifying those advertisements. The advertising sampling application 654 stores the data retrieved by the Web page retrieval system 322 and processed by the Web browser emulation environment 324, advertisement extractor 326, and the structural classifier 328 in the advertising sampling data

664 state and the database 200. Figures 7A, 7C, 7D, and 7E describe, in greater detail, the process that the advertising sampling application 654 follows for each URL 114, 116, 118 identified by the traffic sampling system 120. Even though Figure 6 depicts the central processor 616 as controlling the advertising sampling application 654, a person skilled in the art will realize that the processing performed by the advertising sampling application 654 can be distributed to a separate system configured similarly to the advertising prevalence system 130.

Replace the paragraph on page 44, line 6 through page 44, line 15 with the following:

After the traffic analysis application 652 and the advertisement sampling system 654 process the URL 114, 116, 118 identified by the traffic sampling system 120, the visit object 642 invokes a method in the statistical summarization application 656 to compute summary statistics for the data. The statistical summarization application 656 computes the advertising impression, spending, and valuation statistics for each advertisement embedded in URL 114, 116, 118. The statistical summarization application 656 stores the statistical data in the statistical summarization data 666 state and the database 200. Figure 7F describes, in greater detail, the process that the statistical summarization application 656 follows for each URL 114, 116, 118 identified by the traffic sampling system 120. Even though Figure 6 depicts the central processor 616 as controlling the statistical summarization application 656, a person skilled in the art realizes that the function performed by the statistical summarization application 656 can be distributed to a separate system configured similarly to the advertising prevalence system 130.

Replace the paragraph on page 44, line 16 through page 45, line 6 with the following:

Figure 7A is a flow diagram of a process in the advertising prevalence system 130 that measures the value of online advertisements by tracking and comparing online advertising activity across all major industries, channels, advertising formats, and types. Process 700 begins, at step 710, by sampling traffic data from the Internet 100. Figure 7B describes step 710 in greater detail. Step 720 uses the sampled traffic data from step 710 to perform site selection, and define and refine site definitions for the advertising prevalence system 130. Step 730 uses the result of the site selection and definition process to generate a probe map based on the sampled traffic data. Figure 7C describes step 730 in greater detail. Step 740 uses the probe map from step 730 to visit the Internet 100 to gather sample data from the probe sites identified in step 730. Figure 7D describes step 740 in greater detail. For each URL retrieved in step 740, step 750 extracts the advertisements from the URL, step 760 classifies each advertisement, and step 770 calculates the statistics for each advertisement. Figures 7E and 7F describe, respectively, steps 760 and 770 in greater detail. Finally, process 700 performs data integrity checks in step 780 to verify the integrity of the data and analysis results in the system.

Replace the paragraph on page 45, line 7 through page 45, line 14 with the following:

Figure 7B is a flow diagram that describes, in greater detail, the process of sampling traffic data from Figure 7A, step 710. Process 710 begins in step 711 by gathering data from a Web traffic monitor such as the traffic sampling system 120. Process 710 strips the user information from the data retrieved by the Web traffic monitor in step 712 to cleanse the data and guarantee the anonymity

of the sample. For each URL in the cleansed sample, step 713 measures the number of Web page views observed in the traffic data. Step 714 completes process 710 by statistically extrapolating the measured number of Web page views in the sample to whole universe of the Internet 100.

Replace the paragraph on page 45, line 15 through page 45, line 22 with the following:

Figure 7C is a flow diagram that describes, in greater detail, the process of generating a probe map based on sampled traffic data from Figure 7A, step 730. Process 730 begins in step 731 by analyzing a subset of the sample traffic data that falls within eligible site definitions. Following the analysis in step 731, step 732 builds an initial probe map based on the sample traffic data. Step 733 analyzes the historic advertisement measurement results in the database 200 for the URLs in the initial probe map. Step 734 uses these historic results as well as system parameters to optimize the sampling plan. Step 735 completes process 730 by monitoring the sample results and adjusting the system as necessary.

IN THE CLAIMS

Please amend claims 1-9, 12, 14-15, 17, and 19-53 and add claims 54-69. All of the pending claims in the application, claims 1-69, are reproduced below.

Amended Claims

1. (Amended One Time) A system for estimating prevalence of digital content on a network, comprising:

an estimating device that stores traffic data collected from the network;